Coupled Semiempirical Molecular Orbital and Molecular Mechanics Model (QM/MM) for Organic Molecules in Aqueous Solution

PETER L. CUMMINS, JILL E. GREADY

Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, ACT, 2601 Australia

Received 24 September 1996; accepted 11 March 1997

ABSTRACT: A coupled quantum mechanical and molecular mechanical (QM/MM) model based on the AM1, MNDO, and PM3 semiempirical molecular orbital methods and the TIP3P molecular mechanics model for liquid water is presented. The model was parameterized for each of the three molecular orbital methods using the aqueous solvation free energies of a wide range of neutral organic molecules, many of which are representative of amino acid side chains. The fit to the experimental solvation free energies was achieved by varying the radii in the van der Waals (vdW) terms for interactions between the solute, which was treated quantum mechanically, and the molecular mechanics (TIP3P) solvent molecules. It is assumed that the total free energy can be obtained as the sum of components derived from the electrostatic terms in the Hamiltonian plus a generally smaller ''nonelectrostatic'' term. The electrostatic contributions to the solvation free energies were computed using molecular dynamics (MD) simulation and thermodynamic integration techniques; the nonelectrostatic contributions were taken from the literature. It was found that the experimental free energies could be reproduced accurately (to within 1 kcal/mol) from the MD simulations, provided that the vdW parameter associated with hydrogen bonding (H bonding) was allowed to have different values depending on the QM method (AM1, MNDO, or PM3) and the type of functional group involved in the H bonding. Moreover, the radial distribution functions obtained from the MD simulations using such a parameterization scheme showed the expected H-bonded structures between the solute and molecules of the solvent. The solvent-induced dipole moments also compared favorably with the results of other QM/MM model calculations. *J Comput Chem* **18**: 1496–1512, 1997

Keywords: QM/MM; solvation; free energy; hydrogen bonds; force fields

Correspondence to: J. E. Gready; e-mail: jill.gready@anu.edu.au

Contract/grant sponsor: Australian National University

CCC 0192-8651 / 97 / 121496-17

Introduction

The theoretical study of complex chemical systems consisting of many hundreds or thousands of atoms requires the use of various approximations for the interaction potential between atomic nuclei. These approximations are necessary because the exact quantum mechanical (QM) solution rapidly becomes computationally intractable as the number of atoms is increased. Consequently, the simpler molecular mechanics (MM) potentials are widely used in the simulation of large molecular aggregates and macromolecules to calculate thermodynamic properties or describe phenomena associated with the classical dynamics of the system. However, it becomes essential to use one of the QM methodologies in calculations where, for example, chemical bond formation and breaking or electronic excitation processes are of interest, that is, any process that involves a large change in the electronic structure. Because many of these processes are comparatively localized events (e.g., chemical bond formation), an obvious approach to the computational bottleneck is to treat only a relatively small part of the system quantum mechanically. The remaining larger part of the system would be treated using MM potentials. Singh and Kollman[1] discussed the earliest forms of its implementation and interfaced an *ab initio* molecular orbital program with the AMBER MM/dynamics simulation programs.[2] More recently, Stanton et al.[3–5] experimented with semiempirical and density functional QM interfaces with the AMBER simulation programs. Field et al.[6] incorporated semiempirical QM methods into the CHARMM MM/dynamics simulation program.[7] The combination of coupled QM/MM potentials with general molecular dynamics simulation programs such as AMBER and CHARMM allows the study of chemical processes in complex biomolecular systems.

Any theoretical model to be applied to the study of large systems must necessarily be a compromise between computational efficiency or practicality and the desired chemical accuracy. The main advantage of semiempirical QM methods is that their computational efficiency is orders of magnitude greater than either the density functional or *ab initio* methods. The use of the latter methods limits the types of QM/MM studies that may be conducted, for example, to the electron redistribution within a substrate and cofactor molecule in the active site of an enzyme at a single (X-ray structure) geometry.[8–10] Although useful, such studies do not yield the same amount of definitive information as may be obtained from more computationally demanding investigations on the full reaction pathway using energy gradients. In contrast, a number of QM/MM calculations have now been performed using semiempirical QM methods to compute enzymic reaction mechanisms, potentials of mean force, and solvation energies.[11–18]

Semiempirical quantum mechanics will often provide a reasonable compromise between efficiency and accuracy for the calculation of many properties of isolated molecules, but it usually fails to provide a sufficiently accurate description of the short-range intermolecular interactions, especially hydrogen bonding (H bonding). Although Vasilyev et al.[19] found that QM/MM potentials can provide a better description of H bonding, there remains considerable scope for further refinements of semiempirical QM/MM methods. The models developed by Field et al.[6] and Vasilyev et al.[19] rely heavily on modification of the one-electron terms arising from interaction of the electron cloud of the QM fragments with the point charge of an MM atom, while the parameterizations of the corresponding van der Waals (vdW) terms remains unchanged from the standard MM prescription as found in the CHARMM[7] or OPLS[20] force fields. However, it is well known that the representation of the short-range repulsions is critical for the accurate prediction of the structures and binding energies of H-bonded dimers.[21–24] The size and shape of the solute cavity is similarly important in macroscopic solvation[25] and self-consistent reaction field (SCRF)[26] models. Consequently, in the present work we present a model in which only the vdW terms for interactions between the quantum and molecular mechanics atoms are modified. The one-electron terms require no new parameters but are consistent with the definition of the molecular electrostatic potential (MEP) in the semiempirical approximation,[27] that is, the same as used by Wang and Ford[28,29] and Chudinov et al.[30] in their semiempirical SCRF models. Note that these definitions of the MEP do not employ a density matrix ''deorthogonalization'' procedure.[31] Also, because the MEP is well defined in terms of a wave function, parameterization of the vdW terms would necessarily form the basis of any *ab initio* QM/MM model.[32] Thus, the pro-

posed approach appears to be attractive from a formalistic point of view while retaining a simple functional form for the QM/MM interaction potential.

As for MM potentials and the semiempirical QM methods, QM/MM potentials are only useful if parameterized to reproduce experimental quantities. In the method of Vasilyev et al.[19] the parameters were obtained by fitting calculated H-bond energies to experimental data on ion-molecule complexes in the gas phase. A similar approach was taken by Field et al.,[6] but neutral molecular complexes were also included. Another feature of these two models is that a given parameter is restricted to a common value for all types of complexes. The question arises as to how well these calibrations perform for a wider range of non-H-bonded interactions and for many-body effects that may be encountered in simulations on condensed phases such as in solution or macromolecules. An alternative approach, which would take account of such effects, is to calibrate the model using molecular simulation techniques together with known thermodynamic data. In the present work we adopt such an approach by making use of molecular dynamics (MD) simulation techniques and experimental free energies of solvation. Although this approach requires much larger amounts of computation than for parameter fitting to data on gas-phase bimolecular complexes, it should yield a more accurate model for condensed-phase calculations. Moreover, it should be possible to improve the overall accuracy of the model by allowing the parameter(s) to adopt different values, depending on the type of interaction. We also calculate the solvation free energies using the QM/MM potential of Vasilyev et al.[19] for comparison with the results of the present model.

Methods

INTERACTION MODEL

The total potential energy of a system partitioned into quantum and molecular mechanics groups of atoms may be written in the general form[16]

$$E_T = E_{QM} + E_{MM} + E_{QM/MM}, \qquad (1)$$

where E_{QM} is the energy of the quantum system, E_{MM} is the energy of the molecular mechanics part of the system, and $E_{QM/MM}$ is the interaction energy between the quantum and molecular mechanics parts of the system. In the present model E_{QM} is computed using the semiempirical AM1,[33,34] MNDO,[35,36] and PM3[37] parameterizations. As usual,[16] we write the QM/MM interaction energy as the sum of polar (electrostatic) and nonpolar (vdW) terms,

$$E_{QM/MM} = (1 - \lambda) E_{ele} + E_{vdW}, \qquad (2)$$

where we have introduced the parameter λ for scaling the electrostatic term in the MD simulations in order to determine the solvation free energy. Within the neglect of diatomic differential overlap (NDDO) approximation,[38] the total QM/MM interaction energy may be calculated by the summation of pair-potential energies; that is, the electrostatic term may be written in the form

$$E_{ele} = \sum_{ij} q_i V_j(R_i), \qquad (3)$$

where $V_j(R_i)$ is the electrostatic potential at the position of MM atom i produced by an atom j of the QM solute molecule. The E_{vdW} term is expressed simply as the sum over the usual pair-potential functions from the MM part of the force field.

The charge distribution of the quantum system is polarized by the electrostatic potential produced by the atom-centered charges q_i of the MM system by performing the self-consistent field (SCF) calculation using the appropriately perturbed one-electron Fock matrix elements,

$$F_{\mu\nu} = F^{\circ}_{\mu\nu} + (1 - \lambda)\,\delta \sum_i q_i V_{\mu\nu}, \qquad (4)$$

where $F^{\circ}_{\mu\nu}$ is the matrix element for the unperturbed (i.e., gas phase) system, $V_{\mu\nu}$ is a matrix representation of the potential V_j in eq. (3), and $\delta = 1$ if ν and μ are on the same atom and $\delta = 0$ if ν and μ are on different atoms, as required by the NDDO approximation. Expressions for $V_{\mu\nu}$ that yield the correct MEP within the semiempirical approximation were given previously.[27] Note that they differ from those used in the other QM/MM potentials.[6,19] The essential difference is in the treatment of the two-center terms[34]

$$\gamma = (\rho_i^{QM} + \rho_0^{MM})^2, \qquad (5)$$

which appear in the expressions for the matrix elements $V_{\mu\nu}$. Field et al.[6] set $\rho_0^{MM} = 0$, while Vasilyev et al.[19] scale γ by a factor of 0.095 for the AM1 and MNDO methods and 0.097 for the PM3 method. Bakowies and Thiel[39] proposed the use of a parameterization method to fit the NDDO terms to reproduce *ab initio* electrostatic potentials. Ford and Yang also proposed a type of *ab initio* fitting scheme for the MEP.[40] Recently, Bash et al.[41] used *ab initio* calculations on solute-solvent clusters to completely reparameterize the AM1 Hamiltonian together with vdW terms for the QM/MM interaction. The expressions we used in the present model can be derived by setting the term γ to zero, that is, $\rho_i^{QM} = \rho_0^{MM} = 0$. Note also that Field et al.[6] retain the empirical scaling expressions for the core-core repulsion energy, whereas the present model and the Vasilyev et al.[19] model do not.

SOLVATION FREE ENERGIES

We begin by adopting features of a macroscopic solvent model in which the free energy is calculated as a sum of polar (electrostatic) and nonpolar terms.[25] Whereas Sitkoff et al.[25] used a finite-difference Poisson-Boltzmann equation (FDPB) approach, in the present scheme the electrostatic contribution to the sum is calculated in an MD simulation by the creation or annihilation of the solute-solvent interaction terms via the λ coupling parameter in eqs. (2) and (4). This would correspond to either a charging (solvation) or discharging (desolvation) of the solute in water. In the actual simulations, the systems were first equilibrated with $\lambda = 0$, that is, the full electrostatic interaction. The electrostatic free energy was then calculated as a desolvation term (ΔG_{desol}) using the slow growth formula[42,43]

$$\Delta G_{desol} = -N^{-1} \sum_{i=1}^{N} [E_{ele}]_i, \qquad (6)$$

where N is the number of integration time steps and $[E_{ele}]_i$ is the electrostatic part of the QM/MM interaction [eq. (3)] evaluated using the configuration obtained at the ith time step in the MD simulation. Equation (6) is readily derived from the equations for thermodynamic integration.[43] Note that the integration method is more efficient than the alternative perturbation method,[42] because the latter would require additional SCF calculations in order to form the necessary statistical averages. The remaining nonpolar (i.e., so-called cavity and vdW) free energy term ($\Delta G_{c/vdW}$) assumes a linear relationship between free energy and the solvent-accessible surface area of the solute as used by Sitkoff et al.[25] For a wide range of solutes this term ranges in value from ca. 2 to 3 kcal/mol. For the set of solutes used to parameterize our model the maximum difference is only 0.78 kcal/mol. Thus, this term is well suited for use in a solvation model because it is relatively small and varies little between solutes. Moreover, it is not feasible to compute this free energy term from the simulation data because it depends not only on contributions from the nonbonded (vdW) potential energy terms but also on changes in translational and rotational degrees of freedom on going from the gas to solution phases.

When a molecule is placed in a polar solvent, the strong electric field will cause a distortion of the molecular geometry. Using a Monte Carlo simulation approach, Gao[15] found that changes in the geometry of solute molecules induced by interaction with the solvent changes the free energy by ca. 5%, which we estimate is likely to be of the same order as the uncertainties in the thermodynamic integration. Moreover, we used the TIP3P model[44] for liquid water that assumes a rigid molecular frame. Consequently, the MD simulations were carried out using bond lengths fixed at their AM1, MNDO, or PM3 gas-phase values. All other internal degrees of freedom in the solutes (bond angles and dihedrals) were left unconstrained.

MOLECULAR MECHANICS PARAMETERS

The initial set of vdW parameters for the solutes were selected from the force field of Weiner et al.,[45] the ''all-atom'' AMBER force field. It is well known that such parameterizations of the 6−12 Lennard-Jones terms for the solute-solvent interactions are inadequate for describing hydrogen bonding. To overcome this deficiency, the AMBER force field uses 10−12 potential functions instead of the usual 6−12 terms to describe the H-bonded interactions. Although other potential functions would suffice for our purposes, in the present model we retained the 10−12 potential in the form,

$$V(R) = D_e\{5(R_e/R)^{12} - 6(R_e/R)^{10}\}, \qquad (7)$$

where R_e is the interatomic separation at which $V(R)$ has a minimum value of $-D_e$. Because most of the binding energy between polar molecules is electrostatic in origin,[21–24] the pair dissociation energies (D_e) of the 10–12 potentials were set to the arbitrary small value of 0.01 kcal/mol so that the contribution of these terms to the total potential energy [eq. (1)] was negligible. Note that R_e is not exactly equal to the H-bond distance but marks the distance for the start of the short-range repulsive interactions that, together with the electrostatic/polarization interactions, determine the H-bond geometry. Initial guesses for R_e ranged between 1.00 and 2.00 and simulations were carried out varying R_e by steps of 0.05 Å in order to find values that best fit the experimental solvation free energies. If the experimental free energies could not be reproduced to within 1 kcal/mol by varying R_e alone, selected vdW radii from the initial AMBER set were then reparameterized together with R_e. The radial distribution functions for the solute-solvent interactions were also computed in order to determine whether the model was capable of reproducing the expected H-bonded structures.

MOLECULAR DYNAMICS SIMULATIONS

The solvated systems were initially energy minimized for 100 cycles of conjugate gradients, followed by the MD simulations starting with $\lambda = 0$ in eqs. (2) and (4). After an initial 10 ps of MD, which was sufficient time for equilibration of the solute-solvent interaction energy on which the solvation free energies depend [eq. (6)], the radial distribution functions, solute dipole moments, and the electrostatic components of the solvation free energies were computed. The free energies were obtained by integration between $\lambda = 0$ and $\lambda = 1$ over a 300-ps simulation time with λ coupled continuously and linearly to the MD time step.[43] We used the constant temperature and pressure algorithm of Berendsen et al.[46] to perform the MD simulations with a time step of 0.002 ps and the SHAKE algorithm[47] to constrain all bond lengths to the specified equilibrium values (i.e., gas-phase equilibrium values calculated at the AM1, MNDO, and PM3 levels for the solute molecules). The temperature was set at 300 K with a relaxation time of 0.1 ps and the pressure was set at 1 atm. A 9-Å cutoff was used for the neglect of nonbonded solute-solvent and solvent-solvent interactions. All calculations were performed using Molecular Orbital Programs for Simulations (MOPS).[48]

Results and Discussion

The 6–12 potential vdW radii (R^*) and well depth (ε) parameters obtained are given in Table I, and the R_e values obtained for the 10–12 potential [eq. (7)] are given in Table II. The nonpolar cavity/vdW contributions ($\Delta G_{c/vdW}$), and the total free energies of solvation (ΔG_{sol}) for the AM1, MNDO, PM3, and Vasilyev et al.[19] (AM1/OPLS) models are given in Table III, together with the total free energies obtained from the FDPB approach and experiment.[25] Note that the total free energy was obtained by adding the calculated $\Delta G_{ele} = -\Delta G_{desol}$ [eq. (6)] values to the $\Delta G_{c/vdW}$ values obtained from Sitkoff et al.[25] Using the

TABLE I.
Parameters Obtained for Lennard-Jones 6–12 Potential for Use in QM / MM Models.

Atom[a]	R^* (Å)	ε (kcal / mol)	Description of Atom
H	1.00	0.02	Hydrogen attached to N, O, OH, OW, S
HT	1.54	0.01	Hydrogen attached to CT
C	1.85	0.12	Carbon except *sp*3 and aromatic
CA (AM1)	2.05	0.12	Aromatic ring carbon
CA (MNDO)	1.60	0.12	Aromatic ring carbon
CA (PM3)	1.80	0.12	Aromatic ring carbon
CT	1.80	0.06	Tetrahedral (*sp*3) carbon
N	1.75	0.16	Nitrogen
N (PM3)	1.70	0.16	Amines and indole nitrogen
O	1.60	0.20	Carbonyl oxygen
OH	1.65	0.15	Oxygen attached to hydrogen in alcohols carboxylic acids and phenols
OH (MNDO)	1.60	0.15	Oxygen attached to hydrogen in phenols
OH (PM3)	1.60	0.15	Oxygen attached to hydrogen in phenols
OW	1.768	0.152	Water oxygen (TIP3P)
S	2.0	0.20	Sulfur

[a] Values for the atomic force-field parameters, R^* and ε, are the same for AM1, MNDO, and PM3 unless specific for the QM method given in parentheses.

TABLE II.

Parameters R_e (Å) Obtained for 10 – 12 H-Bond Pair Potentials (with D_e = 0.01 kcal / mol) in AM1, MNDO, and PM3 Models for Different Solute Types and Functional Groups.

Solute Type	Group[a]	H Bonding[b]	AM1	MNDO	PM3
Alcohol	—O—H	OH···H	2.35	1.90	1.90
		H···OW	1.80	1.50	1.50
Amine	>N—H	N···H	2.65	2.40	2.20
		H···OW	1.90	1.80	1.60
Thiol	—S—H	S···H	3.95	3.40	3.95
		H···OW	2.55	2.00	2.55
Sulfide	>S	S···H	3.75	3.30	3.90
Aldehyde and ketone	>C=O	O···H	2.75	2.60	2.70
Carboxylic acid	>C=O	O···H	2.65	2.65	2.65
	—O—H	OH···H	2.35	1.95	1.90
		H···OW	1.80	1.55	1.50
Guanidine	>N=	N···H	—	2.60	2.70
	—NH_2	H···OW	—	2.00	2.05
Primary amide	>C=O	O···H	2.70	2.50	2.60
	—NH_2	H···OW	2.10	2.15	2.10
N-Alkyl-amide	>C=O	O···H	2.55	2.35	2.40
	—NH_2	H···OW	2.15	2.00	2.00
Phenol	—O—H	OH···H	1.90	2.00	2.10
		H···OW	1.80	2.00	1.90
Pyridine	>N=	N···H	2.60	2.60	2.55
Indole	>N—H	H···OW	2.15	—	—
Imidazole	>N=	N···H	2.55	2.60	2.75
	>N—H	H···OW	2.15	2.15	2.10

[a] Proton acceptor / donor group of the solute.
[b] See Table I for definition of atom types.

parameterization schemes in Tables I and II, very good agreement can be obtained between the AM1, MNDO, and PM3 simulation free energies of solvation and the FDPB and experimental results. The AM1, MNDO, and PM3 models give root mean square (RMS) deviations from experiment of ca. 0.3 kcal/mol.

To test the quality of the sampling in the calculation of ΔG_{ele}, we also carried out the reverse integration (i.e., from zero electrostatic interaction to full electrostatic interaction) for a small number of selected solutes. For methanol (AM1) we obtained an absolute difference of 0.18 kcal/mol, while the difference for *N*-methylacetamide (AM1) was only 0.15 kcal/mol. These differences represent a mere few percent of the total solvation free energy and give a great deal of confidence that the integration is proceeding slowly enough to ensure that configuration space is being thoroughly sampled. We successfully used the same techniques to study the relative solvation of molecular ions in previous studies.[43]

The 6–12 Lennard-Jones potential parameters in Table I are mostly the same as in the original Weiner et al.[45] (AMBER) force field. In a few cases the radii for oxygen atoms (OH) and nitrogen (N) were decreased by 0.05 Å. The aromatic carbon atom (CA) radii were increased from 1.85 Å in the AMBER force field to 2.05 Å for the AM1 method and decreased to 1.60 and 1.80 Å for the MNDO and PM3 methods, respectively. Note that the OPLS carbon radii are larger than in the AMBER force field to compensate for the nonpolar hydrogens having effectively zero radii. Hence, satisfactory results for the solvation free energies were also obtained for the aromatics benzene, toluene, and naphthalene using the AM1/OPLS model where R^*(CA) = 2.10 Å.

TABLE III.
Free Energies (kcal / mol) of Solvation Computed Using AM1, MNDO, and PM3 QM / MM Models.

Solute (Amino Acid)[b]	$\Delta G_{c/vdW}$ Ref. 25	ΔG_{sol}[a] AM1	MNDO	PM3	AM1 / OPLS[c]	FDPB[d]	Exp.[e]
Methanol (Ser)	1.77	−4.80	−4.61	−5.03	−2.57	−5.44	−5.08
Ethanol (Thr)	1.95	−4.85	−5.05	−4.93	−2.82	−5.02	−4.90
Propanol	2.11	−4.39	−4.47	−4.94	−2.76	−4.84	−4.83
Isopropyl alcohol	2.09	−5.18	−5.08	−5.41	−2.51	−4.60	−4.76
Ammonia	1.56	−5.31	−5.12	−4.72	−5.16	−4.12	−4.31
Methylamine	1.81	−4.64	−4.45	−4.83	−4.12	−4.96	−4.57
Ethylamine	1.97	−4.63	−4.73	−4.71	−4.14	−4.61	−4.50
Dimethylamine	2.01	−4.44	−4.09	−5.05	−3.22	−4.52	−4.29
N-Butylamine (Lys)	2.30	−4.42	−4.50	−4.38	−3.78	−4.30	−4.38
Methylthiol (Cys)	1.90	−1.20	−1.29	−1.40	−26.43	−1.35	−1.24
Ethylthiol	2.06	−1.30	−1.15	−1.15	−26.84	−1.15	−1.30
Dimethylsulfide	2.07	−1.49	−1.48	−1.29	−22.71	−1.85	−1.54
Methylethylsulfide (Met)	2.24	−1.44	−1.45	−1.55	−23.18	−1.46	−1.49
Acetaldehyde	1.89	−3.40	−3.49	−3.67	−4.38	−3.72	−3.50
Proprionaldehyde	2.06	−3.28	−3.34	−3.30	−4.08	−3.34	−3.44
Acetone	2.08	−3.79	−3.73	−3.85	−5.03	−3.93	−3.85
2-Butanone	2.24	−3.91	−3.74	−3.55	−4.87	−3.56	−3.64
Acetic acid (Asp)	1.97	−6.76	−6.78	−6.80	−6.57	−6.63	−6.70
Propionic acid (Glu)	2.14	−6.34	−6.56	−6.31	−6.43	−6.41	−6.47
N-Propyl-guanidine (Arg)	2.46	−10.32	−11.04	−10.66	−13.02	−10.90	−10.92
Acetamide (Asn)	1.99	−9.15	−8.93	−9.92	−10.27	−9.76	−9.72
Propionamide (Gln)	2.15	−9.08	−9.48	−9.53	−10.33	−9.42	−9.42
N-Methyl-acetamide	2.20	−9.93	−9.64	−10.08	−9.49	−10.00	−10.08
Benzene	2.15	−0.77	−0.49	−0.40	−0.82	−1.06	−0.87
Naphthalene	2.50	−2.50	−2.69	−2.26	−2.75	−2.03	−2.39
Toluene (Phe)	2.33	−0.95	−0.79	−0.89	−1.09	−0.77	−0.76
Ethylbenzene	2.49	−1.06	−0.77	−0.97	−1.08	−0.59	−0.80
Pyridine	2.11	−4.97	−5.13	−4.65	−4.79	−5.06	−4.70
4-Methylpyridine	2.31	−4.76	−4.91	−4.93	−4.81	−5.05	−4.93
Phenol	2.22	−5.85	−6.13	−5.77	−3.94	−6.39	−6.62
p-Cresol (Tyr)	2.41	−5.94	−6.29	−6.18	−4.10	−6.11	−6.13
Methylindole (Trp)	2.55	−6.08	−5.94	−5.66	−5.95	−5.84	−5.91
Methylimidazole (His)	2.19	−10.14	−9.89	−10.15	−8.64	−10.22	−10.25
RMS deviation from exp.		0.32	0.30	0.29	1.27[f]	0.18	

Included for comparison are the AM1 / OPLS, FDPB, and experimental free energies.
[a] $\Delta G_{sol} = \Delta G_{c/vdW} - \Delta G_{desol}$ [see eq. (6)].
[b] Amino acid with corresponding analogous molecular group.
[c] Calculated using the QM / MM potential of Vasilyev et al.[19]
[d] Finite difference Poisson-Boltzmann (FDPB) equation results from ref. 25.
[e] Experimental values as quoted in ref. 25.
[f] Does not include sulfur compounds.

A large number of functional groups were considered for parameterization of the H-bonding terms (Table II). However, note that in some instances the 10−12 potential functions [eq. (7)] were not necessary to reproduce experimental free energies to within the desired accuracy (1 kcal/mol), and consequently some R_e values were not determined. Values for R_e ranged from 1.50 to 2.15 Å for X—H···OW interactions (X = N, O of the solute), i.e., where water is the proton acceptor, and from 1.90 to 2.75 Å for the proton donor X···H interactions. Where a functional group can participate in both types of H bonding (e.g., in methanol), the H bond where water acts as the proton donor

is associated with the larger R_e value. The exception to this seems to be the phenols where the two types of H bonding have similar R_e values. The R_e values for amines are larger than for alcohols, consistent with a larger vdW radius for nitrogen. For alcohols R_e is 2.35 (AM1) or 1.90 Å (MNDO and PM3), whereas for aldehydes and ketones these values increase to 2.75 (AM1), 2.60 (MNDO), and 2.70 Å (PM3). The parameters for carboxylic acids are not very different from those for the corresponding functional groups in alcohols and aldehydes or ketones, although the carbonyl group in amides tends to show greater variations with R_e values ranging from 2.35 to 2.70 Å. The R_e values for sulfur as a proton acceptor range from 3.30 to 3.95 Å, and for sulfur as proton donor we obtain 2.00 (MNDO) and 2.55 Å (AM1 and PM3).

Radial distribution functions, $g(r)$, for the solute-solvent H-bond interactions involving oxygen atoms of the solute are given in Figures 1–3 while those involving nitrogen of the solute are given in Figures 4–6. It is encouraging that although the parameters in Table II were determined from a fit to the experimental free energies of solvation only, Figures 1–6 show that the expected hydrogen bond distances are also reproduced quite well. The $X\cdots H$ distances range between 1.7 and 2.2 Å while $X\cdots OW$ distances are in the 2.8–3.1 Å range. For most of the possible H-bonding sites at least one H bond was predicted. The exceptions are for the $H\cdots OW$ H bond in amines, *N*-propylguanidine, amides (all QM models), and methanol (MNDO), and for the $OH\cdots H$ H bond in acetic acid (MNDO and PM3) where there are ca. 50% probabilities of H bonding. The absence of an $H\cdots OW$ H bond in methylimidazole is likely due to the steric influence of the methyl substituent. However, it is apparent from the present results and from calculations performed on methanol in solution with different MM potentials[49] that the numbers of H bonds are very sensitive to the choice of force-field parameters whereas H-bond distances are much less sensitive. In general the present $g(r)$ functions correlate well with other work.[50,51] In particular, all three QM methods (AM1, MNDO, and PM3) in the present model predict two H bonds between the amide carbonyl group and water, in agreement with other theoretical and experimental studies.[51]

Although the Vasilyev et al. model (AM1/OPLS)[19] was obtained by fitting the calculated H-bond energies to gas-phase data on ion-molecule complexes, in some cases we obtained satisfactory results for the solvation energies of neutral molecules. However, the AM1/OPLS model also gives rise to some very large errors for the free energies of solvation for neutral molecules. In particular, the model breaks down for molecules containing sulfur. The radial distribution functions for the sulfur compounds for the present AM1 model and the AM1/OPLS model are shown in Figure 7. The strength of the electrostatic/polarization interactions between the polar solute atoms and water is greatly overestimated for the AM1/OPLS model, leading to short H-bond lengths and a large number of H-bonded water molecules. In contrast, the present AM1 model predicts only a 50% probability of H bonding. Increasing the value of the parameter that scales γ [eq. (5)] in the AM1/OPLS model would clearly weaken the Coulomb interactions,[38] reduce the solvation free energy, and give much improved agreement with the experiment. These results indicate that using a single parameter value to describe all H-bond types is far too inflexible. In addition, although the carboxylic acid results are in good agreement with experiment, this appears to be partly fortuitous because solvation energies of alcohols are underestimated by ca. 2 kcal/mol and aldehydes and ketones are overestimated by ca. 1 kcal/mol, leading to a cancellation of errors.

Gao and Xia[13] used a Monte Carlo simulation method and the Field et al.[6] AM1 QM/MM model with the TIP3P model for water to compute solvation free energies of organic compounds. Reasonably accurate energies were obtained for many organic compounds containing only oxygen, while errors of up to 5 kcal/mol were obtained for nitrogen containing compounds. In addition, the energy for benzene, and presumably aromatics in general, is underestimated by ca. 3 kcal/mol. The errors are generally very much smaller in the present model, with the error for any molecule tested not exceeding 1 kcal/mol. The solvent-induced dipole moments obtained from the present AM1 calculations and from the Monte Carlo simulations[14] are given in Table IV. The two models appear to give similar dipole moments for solutes in water, although the present model appears to give somewhat larger values.

Conclusions

We have presented a coupled semiempirical (AM1, MNDO, and PM3) and MM (TIP3P) model for organic molecules in aqueous solution. The

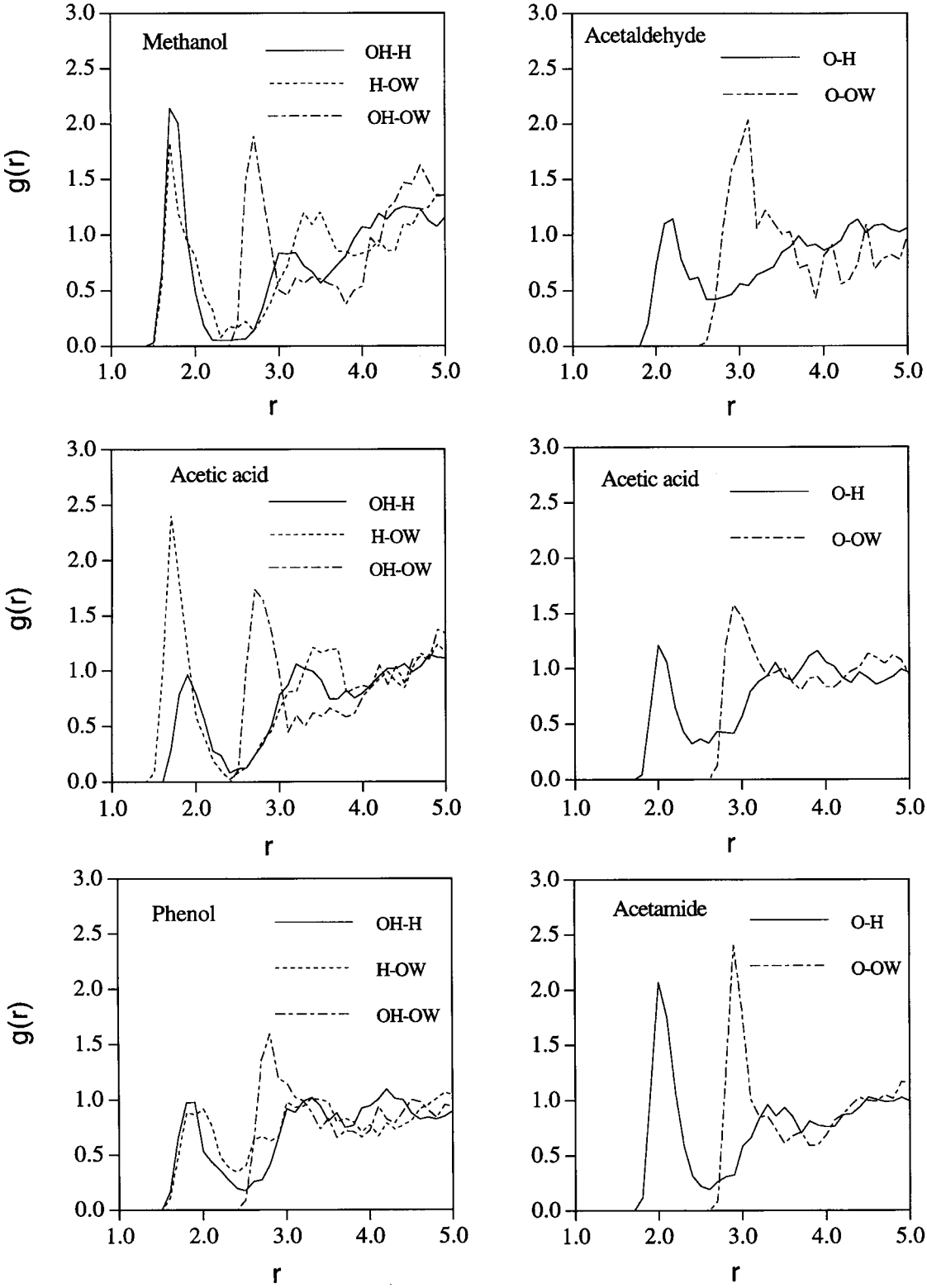

FIGURE 1. AM1 model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving oxygen atoms (see Table I for OH and O) of the solutes.

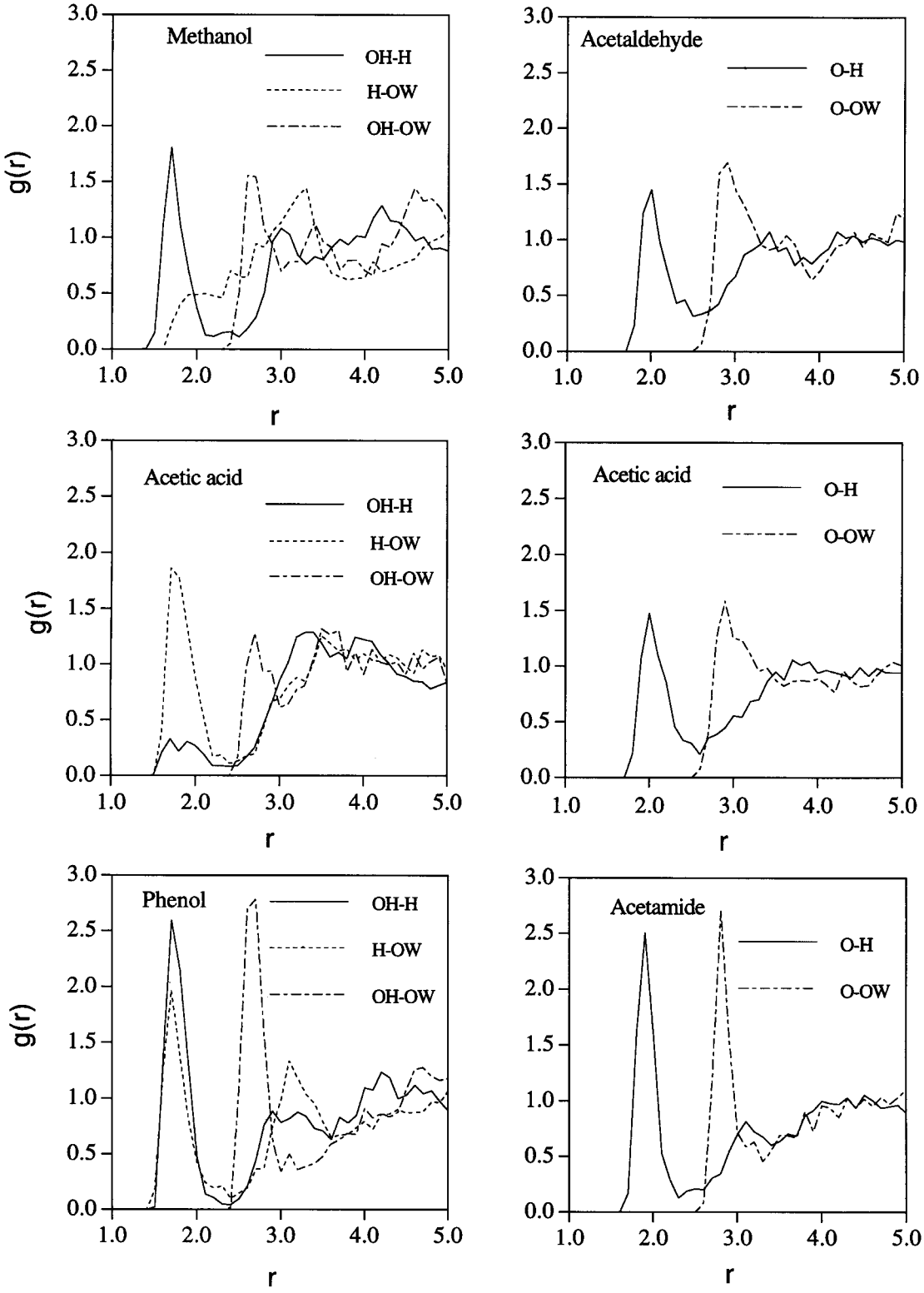

FIGURE 2. MNDO model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving oxygen atoms (see Table I for OH and O) of the solutes.

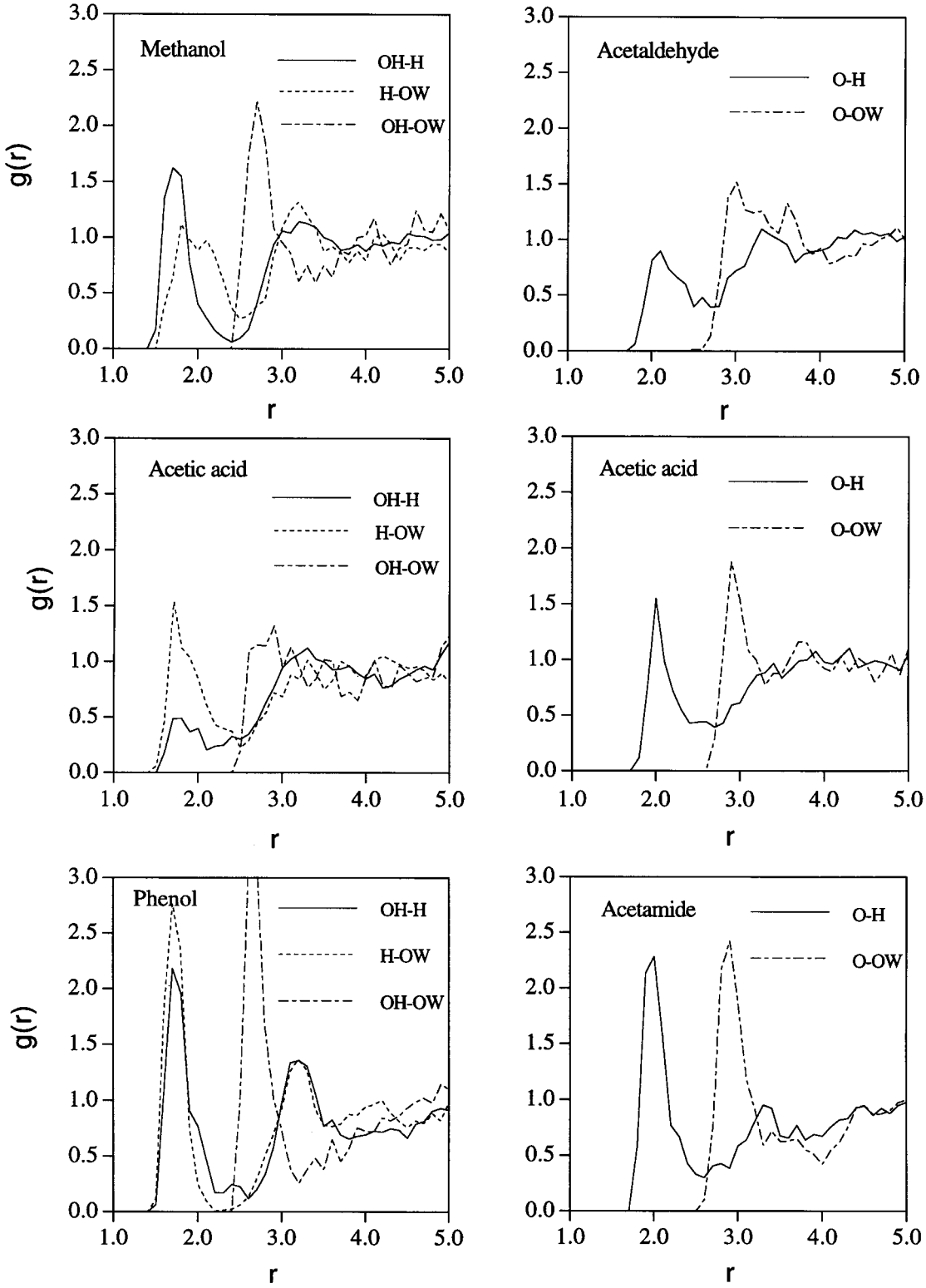

FIGURE 3. PM3 model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving oxygen atoms (see Table I for OH and O) of the solutes.

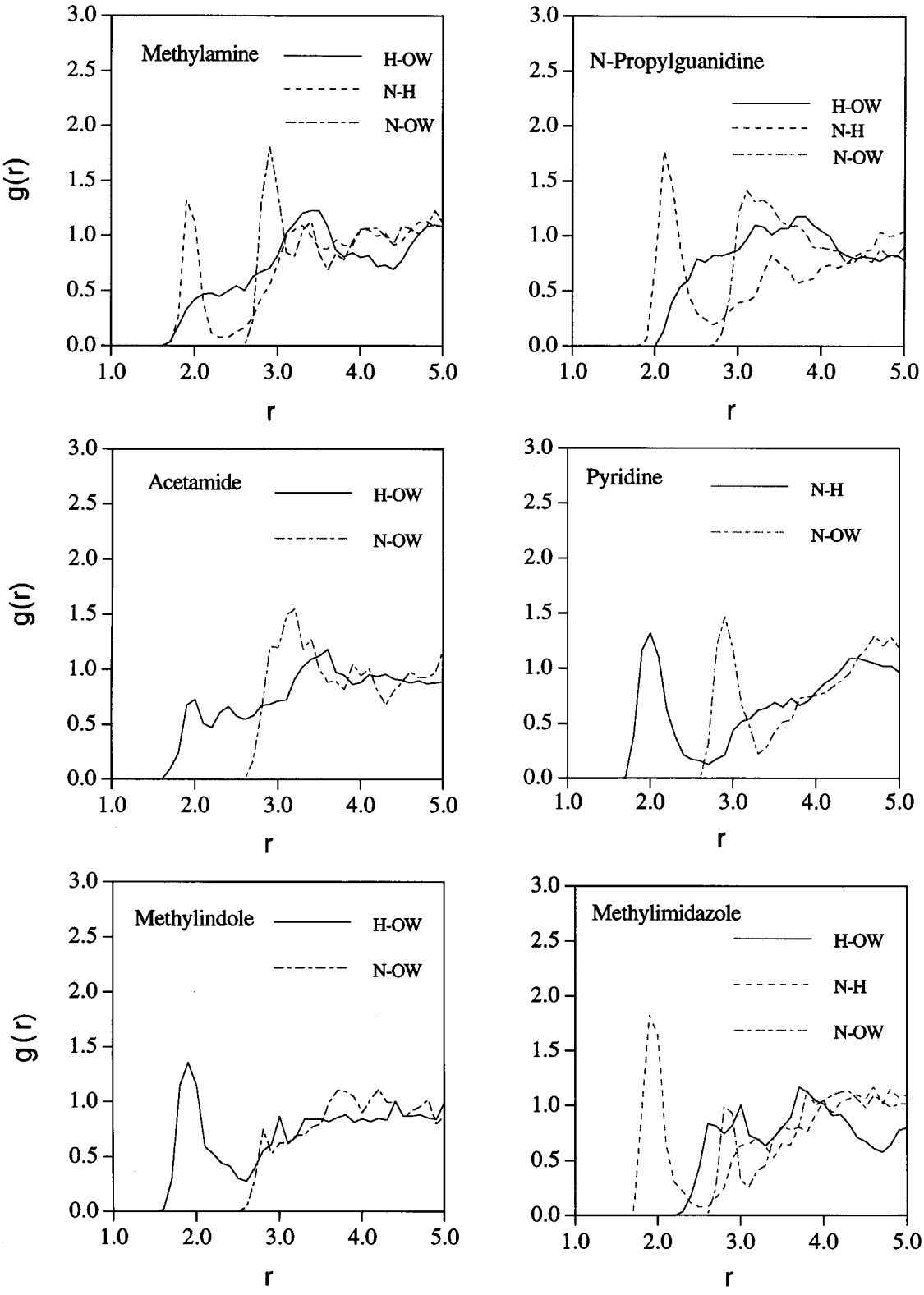

FIGURE 4. AM1 model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving nitrogen atoms of the solutes.

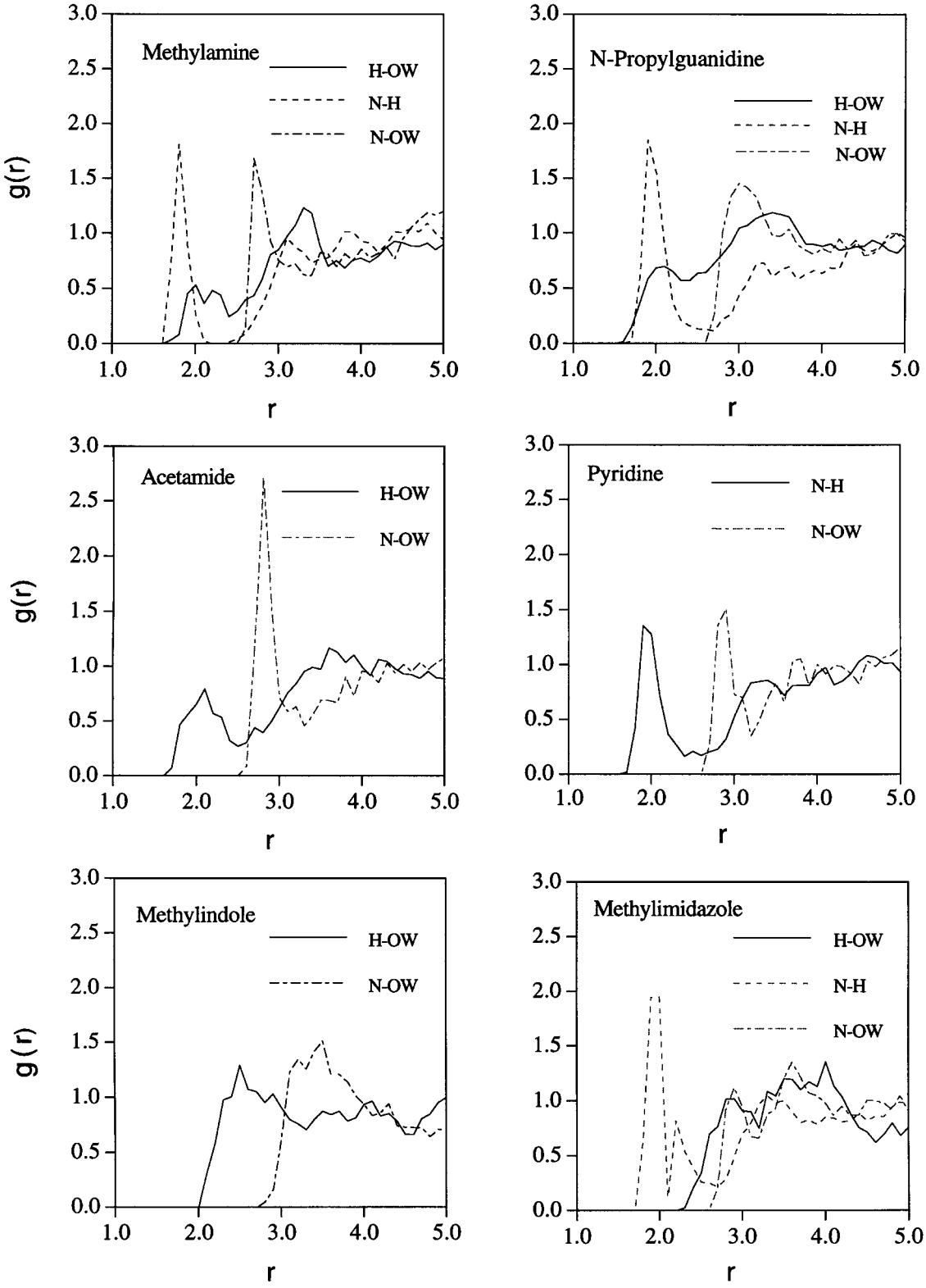

FIGURE 5. MNDO model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving nitrogen atoms of the solutes.

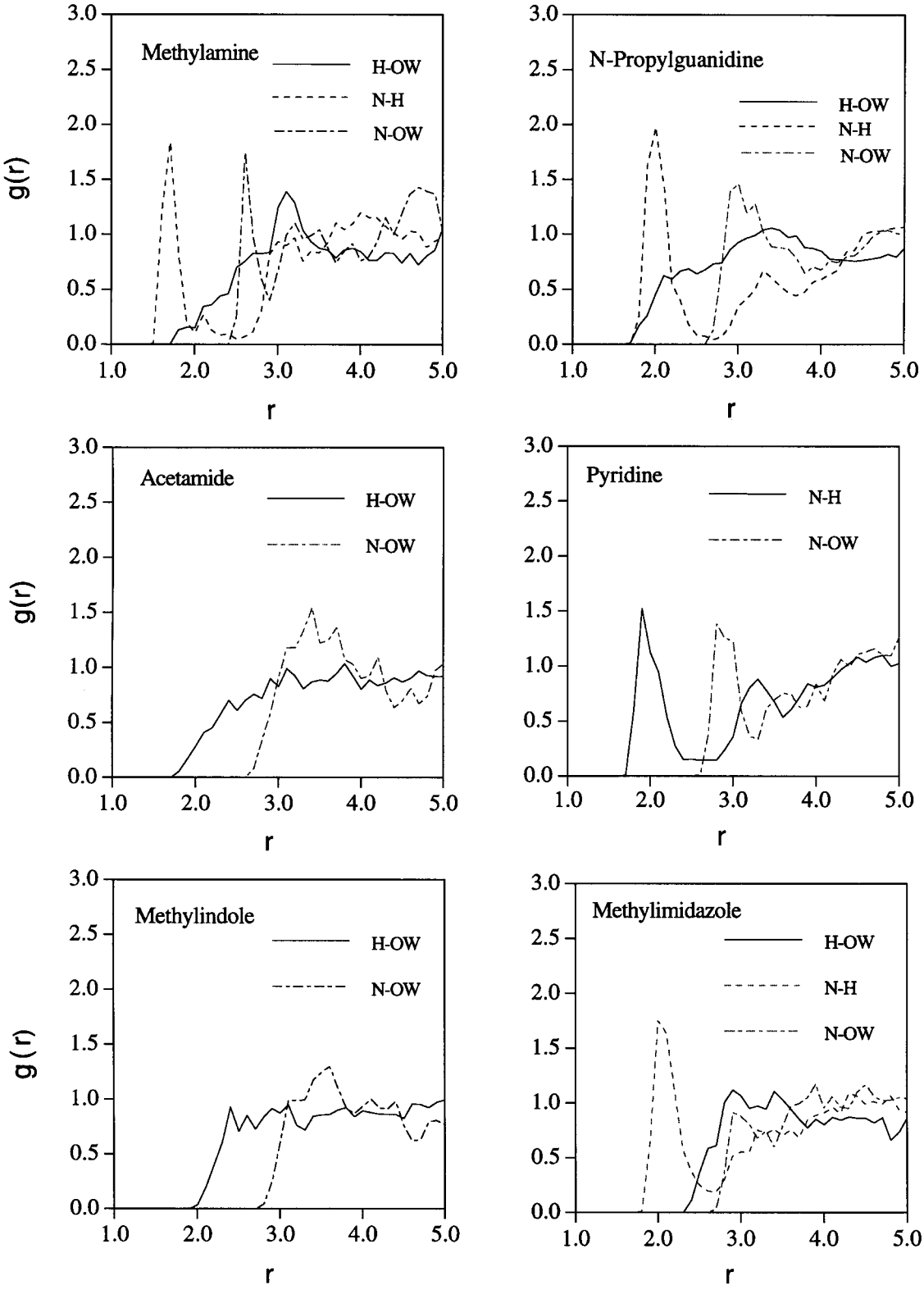

FIGURE 6. PM3 model radial distribution functions $g(r/\text{Å})$ for the solute-solvent H-bond interactions involving nitrogen atoms of the solutes.

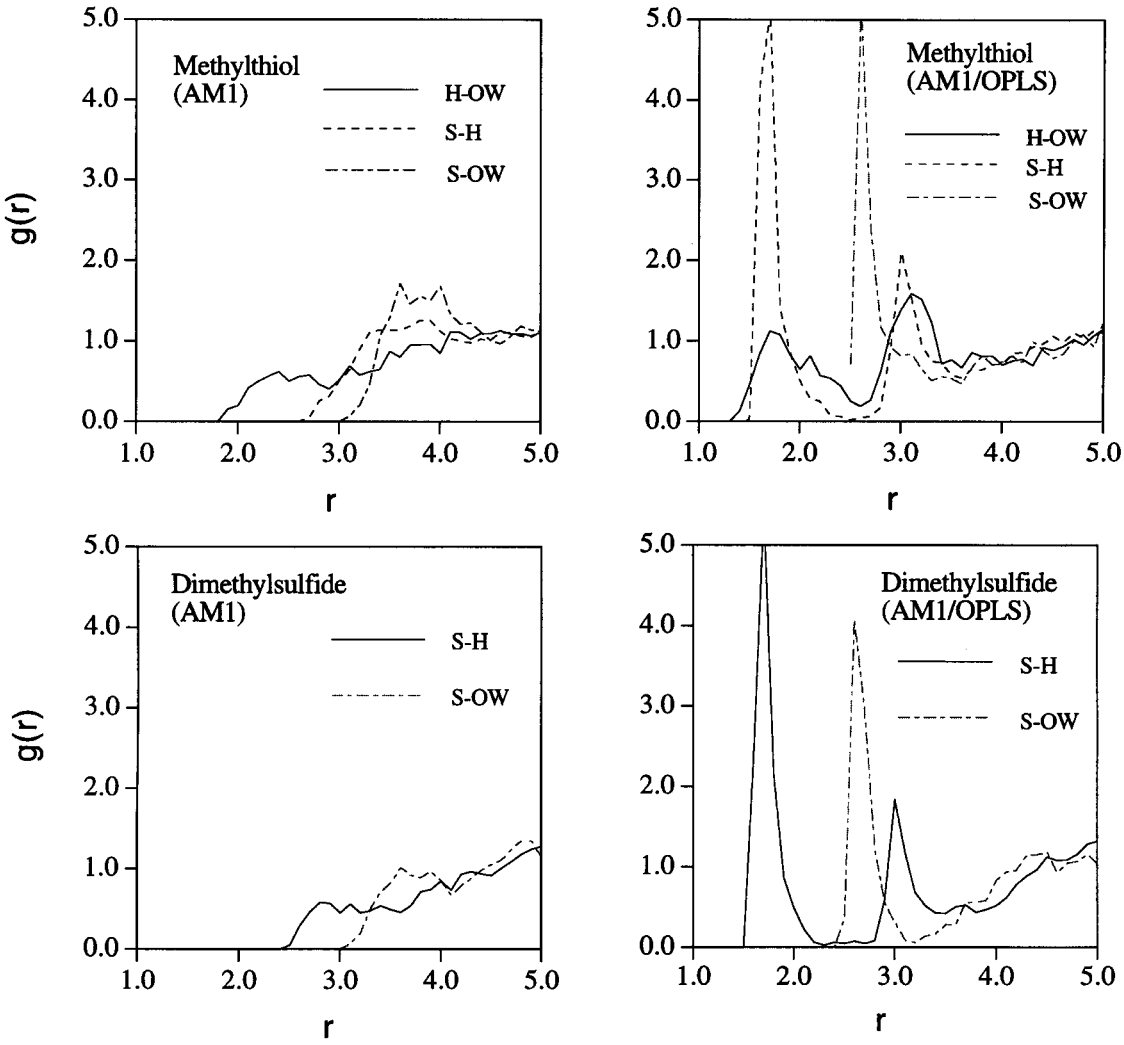

FIGURE 7. AM1 and AM1 / OPLS model radial distribution functions $g(r$ / Å) for the solute-solvent H-bond interactions involving sulfur atoms of the solutes.

TABLE IV.
Gas-Phase AM1 Dipole Moments (μ° / debye) and Solvent-Induced Dipole Moments (μ_{ind} / debye) of Selected Solute Molecules.

Solute	μ°	μ_{ind} (Ref. 14)	μ_{ind} (This Work)
Methanol	1.60	0.41	0.56
Methylamine	1.49	0.20	0.35
Methylthiol	1.76	—	0.42
Acetone	2.92	0.95	0.98
Acetic acid	1.89	0.32	0.46
Phenol	1.23	—	0.97
Pyridine	1.97	—	1.07
Toluene	0.26	—	0.46

parameterization space of the model is restricted to the vdW terms for the interaction between the QM and MM parts of the system, in particular those terms that are likely to have a major influence on hydrogen bonding. For accurate free energies we found that the force-field parameters required different values, depending on the type of functional group involved in H bonding and the QM method (AM1, MNDO, or PM3) used to model the solute molecule. In some instances vdW radii associated with atoms of the solute needed to be changed from the standard MM (AMBER) values. To the extent that other vdW parameter sets differ from the AMBER one that we used in the present

model, the H-bonded terms may need to be reparameterized to retain the same accuracy in ΔG_{sol}. It would also be interesting to see how non-H-bonded interactions are affected by the choice of MM force field. Indeed, a more complete reparameterization of the vdW terms to fit a wider range of data could well be implemented. However, such improvements would come at the expense of increases in the complexity of the fitting procedures.[41]

The solvation free energies computed using the present model are more accurate than those obtained using previously developed QM/MM models[6, 19] that were not specifically parameterized for bulk-phase simulation work. The model is also capable of providing a consistent description of solute-solvent H-bonded structures and solvent-induced dipole moments. Although computationally more expensive than macroscopic models or simulation methods using purely MM potentials, the QM/MM approach has a definite advantage in free energy calculations in that it avoids the problem of the conformational dependence of atomic charges associated with MM force fields. In general, QM/MM potentials may prove to be more accurate than currently available MM potentials.[52] Indeed, the present QM/MM model gives better free energies for amines in solution than can be obtained from calculations performed using standard MM potentials.[53] A particularly important application of the present model would be the study of solvation effects on chemical reactions and in the study of substrate or inhibitor binding to enzymes and catalytic reaction mechanisms.

Acknowledgments

We gratefully acknowledge the Australian National University Supercomputer Facility for generous grants of computer time and the support of an ANU Strategic Development Grant.

References

1. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **7**, 718 (1986).
2. P. K. Weiner and P. A. Kollman, *J. Comput. Chem.*, **2**, 287 (1981).
3. R. V. Stanton, D. S. Hartsough, and K. M. Merz, *J. Phys. Chem.*, **97**, 11868 (1993).
4. R. V. Stanton, D. S. Hartsough, and K. M. Merz, *J. Comput. Chem.*, **16**, 113 (1995).
5. R. V. Stanton, L. R. Little, and K. M. Merz, *J. Phys. Chem.*, **99**, 483 (1995).
6. M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.*, **11**, 700 (1990).
7. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
8. J. Bajorath, J. Kraut, Z. Li, D. H. Kitson, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA*, **88**, 6423 (1991).
9. J. Bajorath, D. H. Kitson, G. Fitzgerald, J. Andzelm, J. Kraut, and A. T. Hagler, *Proteins: Struct. Funct. Genet.*, **9**, 217 (1991).
10. J. Bajorath, Z. Li, G. Fitzgerald, D. H. Kitson, M. Farnum, R. M. Fine, J. Kraut, and A. T. Hagler, *Proteins: Struct. Funct. Genet.*, **11**, 263 (1991).
11. P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, and M. Karplus, *Biochemistry*, **30**, 5826 (1991).
12. V. V. Vasilyev, *J. Mol. Struct. (Theochem.)*, **304**, 129 (1994).
13. J. Gao and X. Xia, *Science*, **258**, 631 (1992).
14. J. Gao, F. J. Luque, and M. Orozco, *J. Chem. Phys.*, **98**, 2975 (1993).
15. J. Gao, *Biophys. Chem.*, **51**, 253 (1994).
16. J. Gao, *Rev. Comput. Chem.*, **7**, 119 (1996).
17. H. Liu, F. Muller-Plathe, and W. F. van Gunsteren, *J. Chem. Phys.*, **102**, 1722 (1995).
18. M. A. Thomson, *J. Am. Chem. Soc.*, **117**, 11341 (1995).
19. V. V. Vasilyev, A. A. Blyinyuk, and A. A. Voityuk, *Int. J. Quantum Chem.*, **44**, 897 (1992).
20. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
21. P. L. Cummins, A. P. L. Rendell, D. J. Swanton, G. B. Bacskay, and N. S. Hush, *Int. Rev. Phys. Chem.*, **5**, 139 (1986).
22. M. A. Spackman, *J. Chem. Phys.*, **85**, 6587 (1986).
23. S.-Y. Liu and C. E. Dykstra, *Chem. Phys.*, **107**, 343 (1986).
24. A. D. Buckingham, P. W. Fowler, and J. M. Hutson, *Chem. Rev.*, **88**, 963 (1988).
25. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.*, **98**, 1978 (1994).
26. E. V. Stefanovich and T. N. Truong, *Chem. Phys. Lett.*, **244**, 65 (1995).
27. P. L. Cummins and J. E. Gready, *Chem. Phys. Lett.*, **225**, 11 (1994).
28. B. Wang and G. P. Ford, *J. Chem. Phys.*, **97**, 4162 (1992).
29. G. P. Ford and B. Wang, *J. Am. Chem. Soc.*, **114**, 10563 (1992).
30. G. E. Chudinov, D. V. Napolov, and M. U. Basilevsky, *Chem. Phys.*, **160**, 41 (1992).
31. F. J. Luque and M. Orozco, *Chem. Phys. Lett.*, 168, 269 (1990).
32. M. Freindorf and J. L. Gao, *J. Comput. Chem.*, **17**, 386 (1996).
33. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
34. M. J. S. Dewar and Y.-C. Yuan, *Inorg. Chem.*, **129**, 3881 (1990).
35. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).
36. M. J. S. Dewar and C. H. Reynolds, *J. Comput. Chem.*, **7**, 140 (1986).

37. J. J. P. Stewart, *J. Comput. Chem.*, **10**, 209 (1989).
38. M. J. S. Dewar and W. Thiel, *Theor. Chim. Acta*, **46**, 89 (1977).
39. D. Bakowies and W. Thiel, *J. Comput. Chem.*, **17**, 87 (1996).
40. G. P. Ford and B. Wang, *J. Comput. Chem.*, **14**, 1101 (1993).
41. P. A. Bash, L. L. Ho, A. D. MacKerell, D. Levine, and P. Hallstrom, *Proc. Natl. Acad. Sci.*, **93**, 3698 (1996).
42. M. Mezei and D. L. Beveridge, *Ann. NY Acad. Sci.*, **1**, 482 (1986).
43. P. L. Cummins and J. E. Gready, *J. Comput. Chem.*, **15**, 704 (1994).
44. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
45. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
46. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
47. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, **34**, 1311 (1977).
48. P. L. Cummins, Molecular Orbital Programs for Simulations (MOPS), 1996.
49. C. Chipot, C. Millot, B. Maigret, and P.A. Kollman, *J. Phys. Chem.*, **98**, 11362 (1994).
50. J. Zeng, J. S. Craw, N. S. Hush, and J. R. Reimers, *Chem. Phys. Lett.*, **206**, 323 (1993).
51. I. P. Gerothanassis, I. N. Demetropoulos, and C. Vakka, *Biopolymers*, **36**, 415 (1995).
52. I. Tunon, M. T. C. Martinscosta, C. Millot, and M. F. Ruizlopez, *Chem. Phys. Lett.*, **241**, 450 (1995).
53. P. Y. Morgantini and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 6057 (1995).